Performance evaluation of the emerging JPEG XT image compression standard

A. Pinheiro [†1], K. Fliegel [*2], P. Korshunov [#3], L. Krasula [*4], M. Bernardo [†5], M. Pereira [†6], and T. Ebrahimi [#7]

[*] *Multimedia Technology Group, CTU in Prague, Czech Republic*
[2] fliegek@fel.cvut.cz, [4] krasuluk@fel.cvut.cz

[#] *Multimedia Signal Processing Group, EPFL, Switzerland*
[3] pavel.korshunov@epfl.ch, [7] touradj.ebrahimi@epfl.ch

[†] *Optics Center and Instituto de Telecomunicações, UBI, Portugal*
[1] pinheiro@ubi.pt, [5] mbernardo@ubi.pt, [6] mpereira@di.ubi.pt

***Abstract*—The upcoming JPEG XT is under development for High Dynamic Range (HDR) image compression. This standard encodes a Low Dynamic Range (LDR) version of the HDR image generated by a Tone-Mapping Operator (TMO) using the conventional JPEG coding as a base layer and encodes the extra HDR information in a residual layer. This paper studies the performance of the three profiles of JPEG XT (referred to as profiles A, B and C) using a test set of six HDR images. Four TMO techniques were used for the base layer image generation to assess the influence of the TMOs on the performance of JPEG XT profiles. Then, the HDR images were coded with different quality levels for the base layer and for the residual layer. The performance of each profile was evaluated using Signal to Noise Ratio (SNR), Feature SIMilarity Index (FSIM), Root Mean Square Error (RMSE), and CIEDE2000 color difference objective metrics. The evaluation results demonstrate that profiles A and B lead to similar saturation of quality at the higher bit rates, while profile C exhibits no saturation. Profiles B and C appear to be more dependent on TMOs used for the base layer compared to profile A.**

I. Introduction

High Dynamic Range (HDR) images are able to represent a wider range of luminance values, closer to the luminance range of the Human Visual System (HVS). For this reason, HDR representation of image and video content is gaining increased interest as a means to improve Quality of Experience (QoE), in particular, in imaging applications such as photography, TV, and Cinema.

HDR images cope with high and low illuminated regions better compared to conventional Low Dynamic Range (LDR) images, which makes HDR images to be more suitable for capturing richer information from the scenes. Typically, HDR images are displayed on legacy monitors using Tone-Mapping Operators (TMOs), which map HDR wider range of contrasts and colors to the ranges available in the displays. JPEG XT is the standard under development that targets specifically the compression of HDR pictures. It uses a base layer, where an LDR image generated by a TMO is compressed keeping the backward compatibility with the conventional JPEG, and a residual layer, where the residual HDR information is coded. A general JPEG XT decoder scheme is presented in Fig.1.

2014 IEEE 16th International Workshop on Multimedia Signal Processing (MMSP), Sep. 22-24, 2014, Jakarta, Indonesia.
978-1-4799-5896-2/14/$31.00

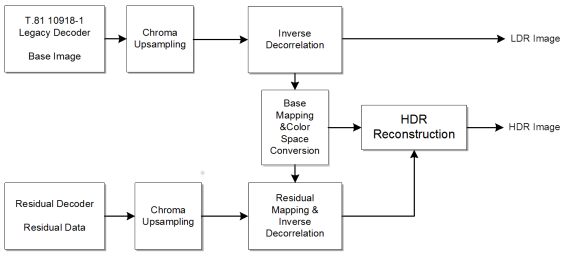

Fig. 1: JPEG XT decoder (from JPEG document wg1n6713-5).

This paper focuses on the objective evaluation of three profiles of JPEG XT. In this paper, we refer to JPEG XT profiles as profiles A, B, and C, in accordance with the way they are identified in JPEG XT specifications. The evaluation is a follow-up from verification tests performed jointly by JPEG HDR Ad Hoc Group and the European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003 - Qualinet).

The evaluations were performed on six HDR images, 'cadik-desk01', 'Knossos8', 'LowerLewisFalls', 'RevelStoke', 'SwissSunset', and 'Zurich2', with first image from Čadík dataset[1] [5] and the others from EMPA dataset[2].

Four TMOs were selected: 'Simple' (a simple linear tone-mapping with inverse gamma correction), 'Reinhard' (by Reinhard *et al.* [1]), 'Drago' (by Drago *et al.* [2]), and 'iCAM06' (by Kuang *et al.* [3]).

The selected HDR images were coded using three profiles with four TMOs used for the base layer. Moreover, different combinations of compression qualities were used for the base layer and for the HDR residual layer coding.

The following are the main objectives of this paper:

- To verify the profiles of JPEG XT by checking if they produce an intended HDR JPEG backward-compatible

[1] http://dcgi.felk.cvut.cz/home/cadikm/tmo/
[2] http://empamedia.ethz.ch/hdrdatabase/index.php

Scene	dynCLog [f-stops]	dyn [f-stops]	Resolution
cadik_desk01	18.86	14.51	1920 × 1080
Knossos8	9.42	9.12	4916 × 3273
LowerLewisFalls	16.39	12.51	3800 × 2516
RevelStoke	9.68	8.62	3846 × 2560
SwissSunset	16.57	13.77	4916 × 3273
Zurich2	13.30	8.66	4916 × 3273

TABLE I: HDR images information.

stream when given an original HDR image and a tone-mapped LDR version.
- To better understand the influence of TMOs on JPEG XT performance.
- To better understand the degree of content dependency in JPEG XT performance.
- To study the influence of parameters q and Q that control the balance of quality and bit rate between LDR and HDR portions of a JPEG XT coded image, and the impact of these parameters on the overall rate distortion performance in two practical scenarios.

After this introduction, the test methodologies will be discussed. Next section describes the selection of the content and the TMOs, as well as the strategy to select appropriate coding parameters in the experiments. This is followed by the results where the impact of content and TMOs on the performance of JPEG XT is shown in two practical scenarios. The paper is concluded with a number of general findings and suggestions for future extensions of this work.

II. Test methodology

A. Content

Six HDR images have been selected for this study, based on their visual content and quality. The dynamic range of the content was also among the selection criteria.

Natural HDR images are often acquired following a process that requires a set of pictures with different exposures from the same scene, which are then fused. If done without special care, HDR images generated by this mechanism tend to have poor edges, or exhibit various motion blur distortions in regions where objects move between exposures. In image selection, any potential edge and blur distortions were carefully examined by observing images on a SIM2 HDR monitor to make sure no such distortions were present. Tone mapped versions of the selected HDR images are shown in Fig. 2. The resolutions and dynamic range of the selected HDR images are provided in Table I. The values in the dynCLog column are the dynamic range in logarithm base 2 considering the absolute maximum and minimum values of each image. The values in the dyn column are the dynamic range in logarithm base 2 considering a more robust maximum and minimum estimation procedure described in [4].

B. Tone-Mapping Operators

The TMOs used to display HDR content on legacy monitors are mostly divided into two groups. Global operators apply the same dynamic range compression function on every pixel in the image, while with local operators, this function varies depending on a neighborhood of the pixels under consideration. An overview and an extensive comparison of existing TMOs was presented by Čadík et al. [5].

In this paper, three global and one local TMOs were selected based on their performance on the selected images and the TMOs are: a 'Simple' TMO, 'Reinhard' global TMO [1], 'Drago' TMO [2], and 'iCAM06' [3]. The 'Simple' TMO first clips a certain percentage of luminance values from both ends of the dynamic range, then linearly scales the remaining values into the interval between 0 and 1, and finally applies the inverse gamma correction. The implementations of 'Reinhard' and 'Drago' TMOs available in Banterle's HDR Toolbox [4] were used. In case of 'iCAM06', the MATLAB code provided by the authors was employed.

The parameters of particular TMOs were set in order to provide visually appealing results. For the 'Drago' TMO and 'iCAM06', the default parameters setting was sufficient, whereas for the rest of the operators, the parameters were optimized according to the method proposed by Krasula et al. [6]. In some cases, the final parameters were additionally manually adjusted to create more natural images. The parameter settings for TMOs can be found in Table II.

C. Coding

JPEG XT encoding depends mainly on two parameters: (i) the parameter q that controls the base layer coding quality index, which is exactly the same parameter that is used in the conventional JPEG compression tools, and (ii) the parameter Q that controls the quality of the residual HDR information.

A large set of the two quality encoding parameters were used for the tests. The base layer parameter q was varied from 40 to 95 with step 5, plus the value of 99. Smaller values than $q = 40$ were not used, since the quality of the base layer for lower q values was considered to be too low for any practical application. As for the residual layer quality parameter Q, it varied from 20 to 95 with step 5, plus 99 value. For the residual encoding, it makes sense to use smaller quality values, since Q essentially controls the compression of the dynamic range of an image. Furthermore, $q = 99$ and $Q = 99$ were added to represent a near lossless example in all JPEG XT profiles under study. However, because of the inherent characteristics of profiles A and B, near lossless representations were not achieved even with these parameters. Only profile C, which combines two layers information in an additive manner, could reach near lossless performance.

In this study, results of two practical use case scenarios are considered. One is a typical JPEG encoding case, when the value of the parameter q is equal to 75, providing tone mapped images with common JPEG compression quality. And another is an optimal case with optimal rate distortion achieved by each profile, when the distortion is estimated by SNR and without any a priori constraint on the quality of LDR image. The latter was obtained by varying both parameters q and Q, and selecting the combinations that provided the best performance after a full search.

cadik_desk01

Knossos8

LowerLewisFalls

RevelStoke

SwissSunset

Zurich2

Fig. 2: Tone mapped versions of the selected HDR images.

Scene	Simple		Reinhard [1]		Drago [2]		iCAM06 [3]		
	clipped %	γ	p_α	p_{white}	Ld_{max}	bias	L_{max}	p	γ
cadik_desk01	3.39	3.6162	0.6470	111.2977	100	0.85	20000	0.7000	1
Knossos8	0.00	3.2710	0.7423	18.0095	100	0.85	20000	0.7000	1
LowerLewisFalls	2.50	2.2000	0.3809	28.6045	100	0.85	20000	0.7000	1
RevelStoke	0.19	2.9150	0.8705	24.7873	100	0.85	20000	0.7000	1
SwissSunset	0.00	4.5830	0.4948	52.7505	100	0.85	20000	0.4617	1
Zurich2	2.15	2.4479	1.0039	5.9081	100	0.85	20000	0.7000	1

TABLE II: Parameters settings of the selected TMOs for particular scenes.

D. Evaluation

SNR, RMSE [4], FSIM [7], and CIEDE2000 color difference [8] metrics were computed for different bit rates and images resulting from the use of different quality parameters. SNR is given by,

$$SNR_{dB} = 10 \log_{10} \frac{P_{image}}{P_{noise}} \tag{1}$$

where P_{image} is the power of the reference image (computed over the three color components) and P_{noise} is the power of the distortion due to compression.

RMSE [4] measures the difference between the original image and the decoded image. However, the results from using RMSE metric did not add any new information when compared to using SNR, hence RMSE metric is not reported in this paper due to the lack of space.

Feature SIMilarity Index (FSIM) [7] is a perceptual metric that results from the Structural SIMilarity index (SSIM) [9]. FSIM adds the comparison of the low-level feature sets between the reference image and the distorted image [7]. Hence, the FSIM analyzes the high Phase Congruency (PC) extracting highly informative features and the Gradient Magnitude (GM), to encode the contrast information. This analysis is complementary and reflects different aspects of the HVS in assessing the local quality of an image. The disadvantage of this metric is that it saturates very quickly for the operating points selected in our experiments. However, this saturation may also be related to the fact that the images perceived quality also reaches a saturation for higher bit rates when observed by human subjects.

Finally, another perceptual measure, developed for color error computation, the CIEDE2000 [8], was used. This color error measure includes not only weighting factors for lightness, chroma, and hue, but also factors to handle the relationship between chroma and hue. The CIEDE2000 computation is not reliable in all color spaces. However, a conversion of the float RGB HDR representation to the CIELAB color space can be performed to allow its computation. This metric showed unreliable and erratic behavior in our experiments and, therefore, is not reported in the paper.

III. Results

The rate-distortion (SNR versus overall bit rate) curves were calculated for the three profiles (A, B, and C), four TMOs ('Simple', 'Reinhard', 'Drago', and 'iCAM06'), and six images ('cadik-desk01', 'Knossos8', 'LowerLewisFalls', 'RevelStoke', 'SwissSunset', and 'Zurich2'). A significant number of operating points were produced by varying the base layer parameter q and the residual layer parameter Q as specified in Section II.

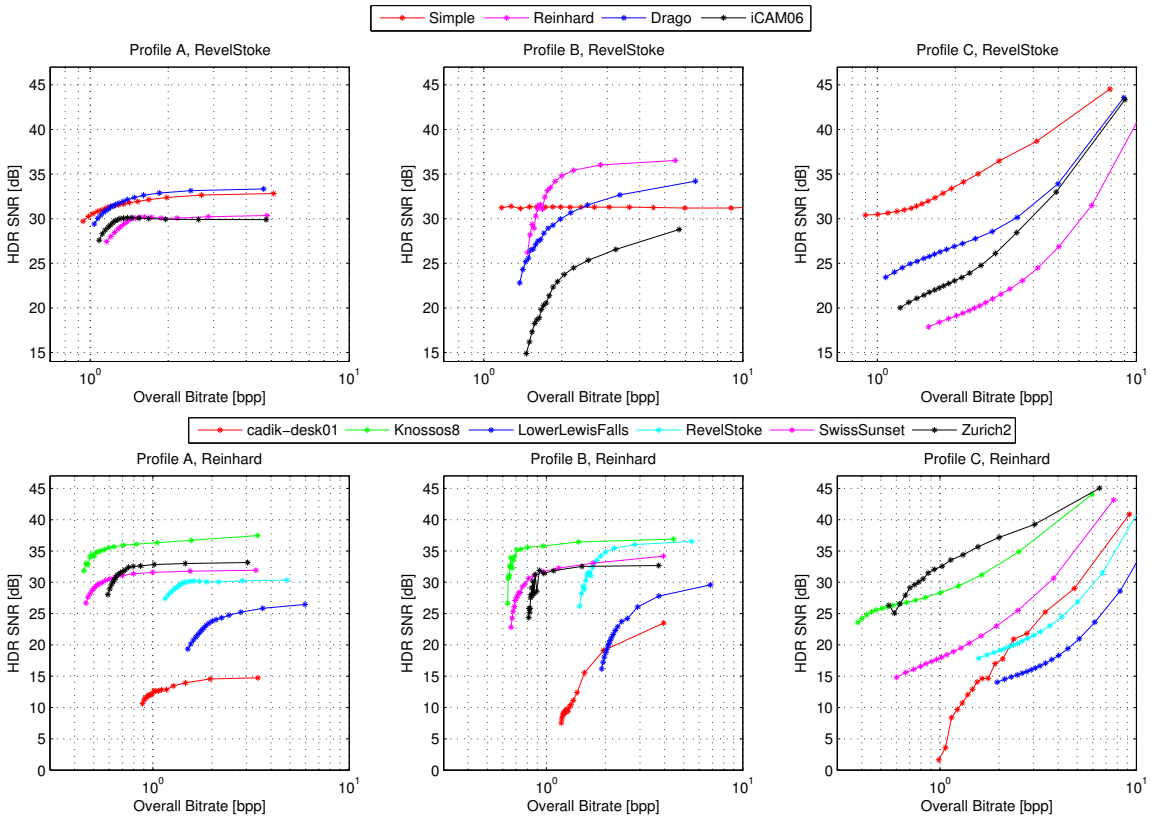

Fig. 3: SNR metric values (varying Q values for q=75) for three profiles of JPEG XT.

To provide performance assessments under realistic situations, in a first use case, the LDR base layer quality was constrained to $q = 75$, since it is a common quality parameter used in JPEG compression. A second use case considers the optimal combinations of q and Q that guarantee the best relation between the SNR and bit rate. No constraints were imposed on the quality of the base layer of the JPEG XT coded image in the second use case.

The paper focuses on the following performance comparative scenarios: (i) four TMOs are used for the same image ('RevelStoke'), and (ii) all six images are encoded while the same TMO is used ('Reinhard'). Figure 3 presents both scenarios for SNR metric when the base layer quality parameter is set to $q = 75$, while Fig. 4 presents both scenarios for SNR metric when the optimal combinations of two quality parameters was used. All plots in both figures have a logarithmic scale in their horizontal axis.

From Fig. 3 the difference in the profiles behavior can be noted. Profiles A and B behave similarly and it can be explained by the intrinsic nature of their encoding mechanisms. Profile C outperforms profiles A and B with higher SNR for the higher bit rates, as the SNR values grow with the increasing bit rate. Implementations of profiles A and B tend to saturate the SNR, while profile C implementation exhibits an increasing SNR as the overall bit rate grows. It may be due to the additive nature of the profile C, since its residual layer is added to the base layer, providing a continuous mechanism to improve the proximity to the base layer data. However, this approach also leads to worse SNR values for lower bit rates. From Fig. 3, profile A outperforms other profiles for the lower bit rates. However, it reaches the saturation quicker than others in the higher bit rates.

The influence of the TMO in the rate-distortion characteristics of the different profiles can be observed in Fig. 3 as well. Different profiles present different rate-distortion curves for different TMOs. Some TMOs can lead to the best performance of one profile and to the worst performance of the other profile. For instance, 'Reinhard' TMO performs the best in profile B and worst in the two other profiles. Figure 3 also suggests that profile A is the least affected by the influence of different TMOs. Profile B, for the 'Simple' TMO and for a fixed q keeps the SNR stable for the increasing Q. This was noticed for all images, except for 'LowerLewisFalls', which passed through an automatic gamma correction, while the others had a fixed gamma correction.

Figure 3 also indicates that image content has an important influence on the SNR in all profiles. This is expected since all profiles compress the TMO versions in the base layer, keeping the backward compatibility with the conventional JPEG decoders. All profiles show a similar behavior considering the content dependency. This means that the most demanding images, in terms of the relation SNR vs. bit rates, are the same for all profiles. The images 'LowerLewisFalls' and 'cadik-desk01' exhibit the lowest SNR values. The similarity between

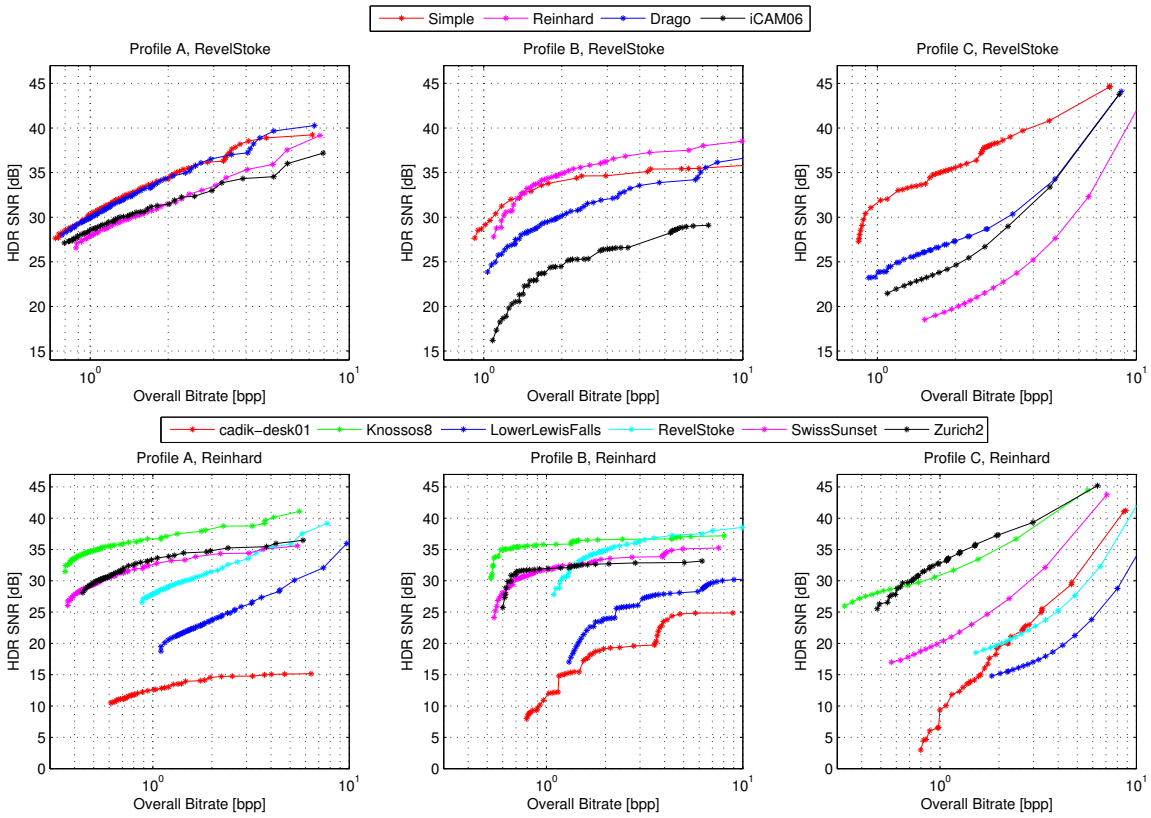

Fig. 4: SNR metric values (combination of Q and q maximizing SNR) for three profiles of JPEG XT.

profiles A and B is evident also in this domain as they clearly separate the images into two different groups of performance. Hence, 'LowerLewisFalls' and 'cadik-desk01' are in the lower performance group, while the other images are in the higher performance group. For the image 'cadik-desk01' all profiles show a poor SNR for the lower bit rates. Moreover, profile A always shows the lowest performance for this image. This image has the highest dynamic range as can be seen in Table I. Finally, it is also important to note that the SNR versus overall bit rate curve of 'Zurich' image in Fig. 3 for profile B presents an irregular behavior for the lowest bit rates considered in this study.

Figure 4 demonstrate results obtained by using the optimal combination of the two quality parameters, q and Q. It can be observed that some curves in the figure exhibit an abrupt change in their tangential behavior. This is due to the discrete nature of the selected operating points that have a fixed step and also a limited range of values. Hence, this figure represents the optimal combinations for the set of computed operating points. This also confirms most of the observations pointed out in Fig. 3. SNR metric saturates for profiles A and B. Moreover, it saturates faster for profile B than for profile A. For the tested values and images, profile C does not present this saturation effect.

Furthermore, from Fig. 4, it can be observed that profile A is the least dependent on the TMO. On the contrary, profile C seems to be the most dependent on the TMO. These plots also show the content dependency of the profiles performance. Profile B reveals the lowest dependency while profile C the highest. For profile A, the SNR saturates very fast for the image 'cadik-desk01'. Furthermore, the profile A reveals to be the most sensitive to the base layer parameter q, while the residual layer Q parameter has a higher influence on profile C. Moreover, profile B is sensitive to both, in the sense that a change in any is always reflected in the relation between bit rate versus SNR.

The FSIM versus bit rate curves are also shown for the three profiles (A, B, and C) in Fig. 5. The figure presents the performance comparison for the six images using the 'Reinhard' TMO with varying Q parameter and fixed $q = 75$. As already discussed in Section II, FSIM metric reaches the saturation very quickly. Profile C is the profile with the quickest saturation, which is the opposite behavior compared to its performance of the SNR metric. The lowest performance denoted in SNR evaluation for 'cadik-desk01' image (see Fig. 3 and Fig. 4) is not observed in the case of FSIM. Otherwise and generally, this metric corroborates the SNR results. The 'Knossos8' image leads to the highest FSIM, which is inline with the SNR case observations. Furthermore the FSIM also confirms the close behavior between profiles A and B observed with the SNR analysis. The FSIM is a perceptual metric and this lack of discrimination could be due to the fact that high bit rates lead to a significantly higher perceived quality for all profiles.

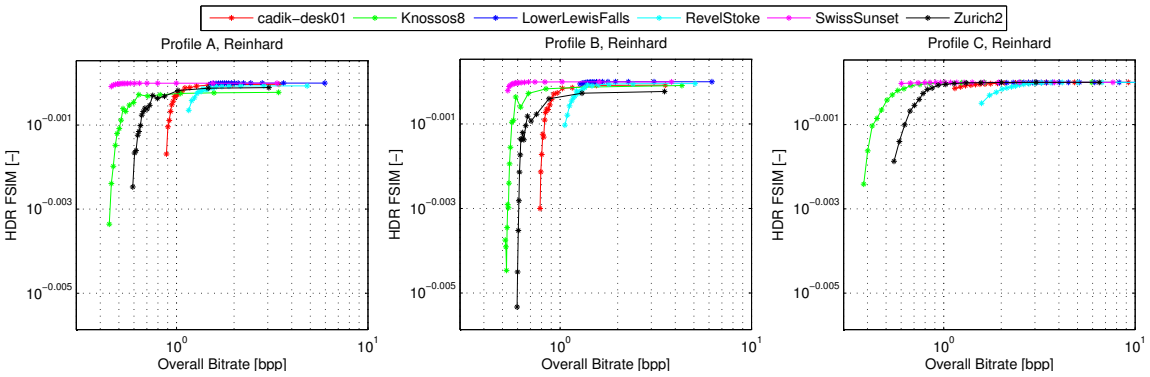

Fig. 5: FSIM metric values (varying Q values for q=75) for three profiles of JPEG XT.

IV. Conclusion

This paper objectively investigates the performance of the three profiles of the upcoming JPEG XT standard for JPEG backward-compatible HDR image compression. Two out of four calculated metrics are reported in the paper: (i) SNR as it is analogous to the popular PSNR metric for LDR images, and (ii) FSIM to represent the class of perceptual metrics. Six images of different dynamic ranges and covering different types of scenes were compressed with each of the JPEG XT profiles using various combinations of parameters q and Q for the base and residual layers of JPEG XT. Six differently tone-mapped LDR versions were also used as the base layer of JPEG XT.

The evaluation results verify the performance of all profiles to be adequate for compression of HDR images. The results for the SNR metric demonstrate that rate-distortion functions of profiles A and B are similar, whereas rate-distortion behaves differently for profile C. It is also evident that Profile A exhibits less dependency on the TMOs used for base layer; profile B shows faster saturation for higher bit rates; profile C, while demonstrating no saturation and is able to encode images with high bit rates, performs worse on low bit rates when compared to profiles A and B. The results using the FSIM metric show faster saturation for all profiles when compared with the SNR. Nevertheless, the profiles shows similarly consistent behavior for the FSIM metric and for the SNR metric.

Since this paper is, to the best of the authors knowledge, the first study that evaluates the performance of all profiles of JPEG XT, more in depth analysis and more metrics are required to draw a more complete picture about JPEG XT performance. Subjective experiments on an HDR monitor are also necessary to fully comprehend how JPEG XT compression affects the perceptual image quality.

Acknowledgments

This work was partially supported by the following projects: COST Action IC1003 European Network on Quality of Experience in Multimedia Systems and Services – Qualinet; EC funded FP7 Network of Excellence VideoSense; Swiss State Secretariat for Education and Research COST research project Compression and Evaluation of High Dynamic Range Image and Video; COST CZ LD12018 Modeling and verification of methods for Quality of Experience (QoE) assessment in multimedia systems – MOVERIQ, Grant No. P102/10/1320 Research modeling of advanced methods of image quality evaluation of the Czech Science Foundation; FEDER funds through the "Programa Operacional Factores de Competitividade – COMPETE" and by Portuguese funds through "FCT – Fundação para a Ciência e a Tecnologia" in the framework of the project PTDC/EIA-EIA/119004/2010 and PEst-OE/EEI/LA0008/2013 and the project PEst-OE-FIS/UI0524/2014.

Moreover, the authors would like to acknowledge Dolby, Trellis and Univ. of Stuttgart for providing implementations for JPEG XT profiles and for the support given during the experiments. Finally, acknowledgments to the Empa HDR Image Database and the Cadik Database for the availability of the HDR images.

References

[1] E. Reinhard, M. Stark, P. Shirley, and J. Fewerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 267–276, 2002.

[2] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," in *Proc. of EUROGRAPHICS 2003*, ser. Computer Graphics Forum, P. Brunet and D. W. Fellner, Eds., vol. 22, no. 3. Granada, Spain: Blackwell, 2003, pp. 419–426.

[3] J. Kuang, G. M. Johnson, and M. D. Fairchild, "iCAM06: A refined image appearance model for HDR image rendering," *Journal of Visual Communication and Image Representation*, vol. 18, pp. 406–414, 2007.

[4] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*. Natick, MA, USA: AK Peters (CRC Press), 2011.

[5] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of HDR tone mapping methods using essential perceptual attributes," *Comput. Graph.*, vol. 32, no. 3, pp. 330–349, 2008.

[6] L. Krasula, M. Narwaria, and P. Le Callet, "An automated approach for tone mapping operator parameter adjustment in security applications," pp. 913 803–913 803–10, 2014.

[7] L. Zhang, D. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[8] M. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Research & Application*, vol. 26, no. 5, pp. 340–350, 2001.

[9] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.